

module_stat4_lecture15

■ Thu, 2/17 4:18PM ⌚ 14:17

SUMMARY KEYWORDS

estimate, module, significant, paper, standard error, blitz, statistical inference, conflict, regression, statistic, coefficient, negative, rainfall, civil, confidence, authors, mu naught, talked, accidents, countries

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:04

So in the last clip, I showed you how to connect regressions with hypothesis testing, and how you can use the tools of hypothesis testing, to check if the regression coefficients are significant or not. In other words, trying to determine if the relationship between one variable and other if that relation is a strong one, is it as a statistically significant connection or not? So before ending the module, I wanted to show you in actual research articles, how to the present the regression estimates, and how do they present if the results are significant or not. So I'm not going to go too deep into the papers. But I'm going to actually show you how, in actual papers, they present results like this, because you mean counter such papers, in different courses during your university life. So let me first look at this particular paper, which is called economic shocks and civil conflict. This is by some economists and political scientists, and it came out in the Journal of political economic in 2004. So what this authors are really interested in, is looking at civil wars. So civil wars have plagued a lot of countries, especially in Sub Saharan Africa. And these authors want to test the hypothesis that does bad economic conditions lead to more chances of civil war. So what we'll focus on is a bunch of sub Saharan African countries, in which the main livelihood of for a lot of people for majority of the people is agriculture. Furthermore, in this country's irrigation systems are fairly poor. So most people depend on rains for their agriculture. So what they're interested in looking at is that does does it happen that in years in which rains are short, right, when rains are lower than normal? So in that case, agriculture in those years agricultural score, economic conditions are poor? And so do we see more civil war in cases like that? In years, where so years where the rainfall is lower than average? Do we see more civil? And is there a significant connection between bad rainfall and more chances of a civil? So what they do is that they run a regression, in which the dependent variable is civil conflict. And the explanatory are the independent variables or growth in rainfall this year, right? So that's your T , and growth in rainfall the previous year. So that's this is t minus one. So what they're running a regression is of this form, $\alpha + \beta_1 x_1 + \beta_2 x_2$. And what they report here are these coefficients on those regresses, right, so the coefficient on x_1 , which is the growth in rainfall this year, that's negative 0.0244. And four rainfall growth in rainfall the previous year, that's negative point one to two. At the same time, what they also represent in these parentheses below those estimates are the standard errors. And furthermore, would they also report using the STAR method, right, so one star means it's significantly different from zero 90% confidence to star means it's significantly

different from zero the 95% confidence, and three star means it's the strongest, that means this coefficient is significantly different from zero at the 99% confidence. So now, if you look at these rates among these two reds, this one here is 95%. With 95% confidence, this is different from zero. So what it means is that remember, the 95% confidence interval is negative 1.96 to 1.96. So what it means is that if you look at the Zed statistic for this, for this estimate, it would be negative point one to two, right divided by the standard error, which is 0.052. So this would take a value for of something like two point something, right. So that means it lies outside this interval. And therefore, it's statistically significant at the 95% level of confidence. But on the other hand, if you look at this one here, that if you divide this by the standard error, you see that that's going to be about something around point five. So that that means this estimate is not significantly different from zero. So translating it into pure ly into just ordinary language, what it would mean is that the growth in rainfall the previous year, right, that has a significant negative impact on civil conflict. So that means if rainfalls were was higher than average the previous year, then that would lower this year's chances of civil conflict. On the other hand, if rainfall was lower than average, the previous year, then it would increase the chances of civil conflict this year. So that's the meaning of this coefficient. Let me present another paper, right. So in if you're discussing such a paper in real course, the professor would spend a lot more time giving you the background of the paper, the background of the data, what these actually mean, in much more depth, right. But what I wanted to just present you how you may actually see results in papers, and at least using the tools that we learned in this course, to try and interpret these estimates. So let me quickly do another paper, this is from a very different area. So this is called Speed Limit enforcement and road safety. By these two authors born Schuster and records. This is a 2019 paper, what they look at is that there are some days this German data. So there are some days in Germany, where police conduct a lot of speed limit enforcement, right, so it's a blitz. And what they want to see is what's the impact on accidents and deaths and injuries due to road accidents on those particular days, compared to other days when there isn't that much enforcement, or there isn't as much of this Blitz? So if you look at the table, right, so what they do is that, so this is the dependent variable, the number of accidents. And they are looking at on this particular day when there is a blitz marathon. So what they're looking at is the impact of this blitz marathon on the number of accidents. And what do you see this is the point estimate. And what it shows is that it's negative. So that means the Blitz marathon decreases the number of accidents. Now, again, the report the standard error below that, and again, use the three star system to denote significance. So what they talk about is the p value, let's see if the P value is less than 0.01. So that means 99% confidence, then they report it by three star, if it's 95% confidence, or P value less than 0.05. They use two star and if there is no star, that means it's not significant. So if you look at this, the number of accidents, right? So this is significantly reduced due to the blitz marathon. What about other variables? Like if we look at the number of fatally injured, right? So it's negative, right? But this is not statistically significant from zero. So that means blitz marathon doesn't have a significant impact on the number of fatally injured. What about those were severe fatal means death, right? severe injury is like your severely injured but not death, right. This also has a negative impact, but this is only significant at the 90% or p less than point. So this is how in actual papers, they represent, report regression results. They also report the standard errors and they report the level of significance. Right? So anything which is nine to 5% or above, that's usually denoted as a strongly significant effect. If it's, if it's below 90%. Right, then it's said to be not significant. Now, let me end the clip and the module with a quick module recap. So we started out by talking about the idea of statistical inference. And the basic idea is this that suppose for some reason, it's impossible or impractical to examine the whole population. So in that case, you look at a part of the population a part of the whole body of people that you're actually interested in finding about it. And then the idea is that how can you look at what's happening in the sample and make inferences about the entire population?

Study statistical inference, we have two main questions. So how do we estimate something? And secondly, how much confidence do we have in RST? So then I talked about the idea of a sample right, and what are the characteristics of an ideal sample. So then, the third thing we talked about is the two main theorems which underlie statistical inference. One is the law of large numbers, which says that as your sample size or n becomes big. So in that case, your estimate comes closer and closer to the population parameter. And secondly, is the central limit theorem, which says that your estimate for large samples is normally distributed. So then, we came to the idea of a point estimate, which could be the mean, or the proportion. And then we talked about standard error, which was, so you have the estimate, and plus minus the error. And then I talked about confidence interval, right. So what's the interval in which I know that 95% of my estimate should lie in and then we talked about hypothesis testing that where the μ is equal to μ_0 versus $\mu \neq \mu_0$ versus $\mu > \mu_0$ or $\mu < \mu_0$. So, this is a two sided test. On the other hand, if you had a one sided test, then it would look something like this $\mu \geq \mu_0$ or $\mu \leq \mu_0$. And in both of these cases, we have the Z-statistic and we compare it. So, that statistic looks like this. And we compare it with the critical value. And the critical value depends on whether we are looking at a two sided test or a one sided test. And then I ended with connecting this hypothesis testing with regressions. So, this ends our module on statistics. And as I've been continuously saying, statistics is a very, very important part of the toolkit of any social scientist. And especially in this age of data and big data, statistics is assuming more and more importance in all our research in all our understanding of social phenomena. So, hopefully you will remain will remember the tools that we learned in this module and be able to apply them in various contexts. So, be sure to work on the practice problems that we have in this module, because they will help deepen your understanding and also help you use the tools that you have learned in this module in actual context.